ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 7 (2013) 370 – 376

Procedia Technology

The 2013 Iberoamerican Conference on Electronics Engineering and Computer Science

Web-based and Interactive Learning - Recognition Method for a Humanoid Robot

Enrique Hidalgo-Peña[a,*], Luis F. Marin-Urias[b], Fernando Montes-Gonzalez[b], Antonio Marin-Hernandez[b], Homero Rios-Figueroa[b]

[a] *Maestría en Inteligencia Artificial*
Sebastián Camacho 5, Centro, Xalapa, Ver. MX 91000
[b] *Departamento de Inteligencia Artificial*
Sebastián Camacho 5, Centro, Xalapa, Ver. MX 91000

Abstract

In this paper an Object Learning and Recognition method for a Humanoid is presented. This method tries to take advantage of the Cloud Resources, since it is based on image web search in order to build training sets for learn about objects appearance. In case of unavailability of Internet access, the robot would ask human to show the object and take the images from its camera. This way, our method aims to be a flexible and natural Human-Robot Interaction framework and to give as much autonomy as possible to the robot.

Keywords: One-class Classification, PCA, Object Recognition, Cloud Robotics, Service Robots

1. Introduction

Nowadays, Cloud resources offer a wide range of possibilities in order to acquire information in almost any topic. For Humans, querying a Search Engine to learn new things has became a common practice. These search engines are able to return results in many ways: documents, images, news, and other media. In other words, there is a huge amount of things the humans can learn using the information available on the Web. This way, it arises the question *"What if a robot makes use of Internet in order to learn how objects look like?"*.

By the other hand, one of the main goals of a potential automated use of Cloud resources is avoiding ambiguity. In the case of this work, and as can be verified empirically, image query results most of times give results that are not visually related to the keyword used for the search. As an example, fig. 1 is presented to show the first results of querying Google for the word "mouse".

In the past few years, the concept of Cloud Robotics has became into a natural framework in which robots make use of the network and Internet resources, as proposed by Arumugam [1] and Kobayashi [2]. These concepts turn very interesting because the capabilities of the robots can be extended and they may

*Corresponding author
Email addresses: ehidalgo@uv.mx (Enrique Hidalgo-Peña), luisfelipe.marin@ieee.org (Luis F. Marin-Urias), fmontes@uv.mx (Fernando Montes-Gonzalez), anmarin@uv.mx (Antonio Marin-Hernandez), hrios@uv.mx (Homero Rios-Figueroa)

Selection and peer-review under responsibility of CIIECC 2013
doi:10.1016/j.protcy.2013.04.046

Fig. 1. Example of ambiguity in search results using "mouse" as keyword.

begin to be more flexible and more powerful than those that just make use of the on board computer's resources.

This work proposes the use of the Google's Image Search Engine, Appearance based Recognition approach and a One-Class Classification algorithm in order to the robot makes recognition process for a particular object, and it does not have more information than the object's name. Main goals on this approach are to give robot as much autonomy as possible and achieve a more natural Human-Robot Interaction.

2. Related Work

In scenarios like real environments constrained light and pose conditions are rarely achieved, so Model-Based approaches like Active Shape Models, presented by Cootes [3] are hard to implement, and in some cases they could be power-processing highly demanding. When handling with service robots, flexibility and time consumption are factors to take into account. Also, when searching for just one object at a time, classic data mining techniques need to be altered in order to manage objects that do not belong to the training set, as we face in this work.

2.1. Feature Extraction for Recognition

Appearance-based systems use Principal Component Analysis (PCA) to extract features from images and also to reduce data dimensionality in the training data-set. Murase and Nayar [4], and Turk and Pentland [5] present this approach.

As proposed by Vicente and Fernández [6], PCA, Independent Component Analysis (ICA) and k-NN classifying algorithm are used for object recognition over the COIL-100 image database. They show in their experiments that symmetric images give better recognition results, as large training data-sets reduce the success rate in classification. On the other hand, they reported that using more components from PCA or ICA increases the success rate.

In [7], Malagon and Fuentes use PCA for pedestrian detection. They proposed using PCA for create four training data-sets for detect pedestrian by taking a new image in grayscale, extracting edges and then projecting each of these on the eigenspace of each of the training data-sets. Then, compare pixel-to-pixel the new images against the training data-sets projections to obtain four differences. Then, a classification value can be established and a classifying criterion is applied.

2.2. One-class classification

There are cases in which training data contains only one class examples, as can be found in the work of Hempstalk [8], El-Yaniv [9], Gesù [10] and Laparra [11]. One-class classification supposes the metric modification to the multi-class algorithms. Such variations are density-based, boundary and reconstruction methods to find a decision threshold.

In [12], Munroe and Madden proposed the use of 1-NNDD for vehicle recognition training just with one type at a time. The motivation of their work rely on the fact that multi-class classification can report unstable results when classifying instances that do not belong to any class in the training set, while this fact does not affect the performance in one-class scenario.

3. Problem Discussion and Approach

As we exposed previously in section I, the goal of this work is to find a method for a service robot to learn an object's appearance analyzing images downloaded from Internet with the object's name as the search keyword. The keyword is the only *a priori* object's information the robot has about it. The main assumption in this scenario is that keyword-based image search returns only one-type object results. However, our assumption is not valid for all keywords, due to ambiguity, thus a challenge in this scenario is that in many cases the search results show images non visually-related to the object's name.

The schematics of our model is presented in figure 2, where is appreciated how robot can learn about objects and recognize them from Web images.

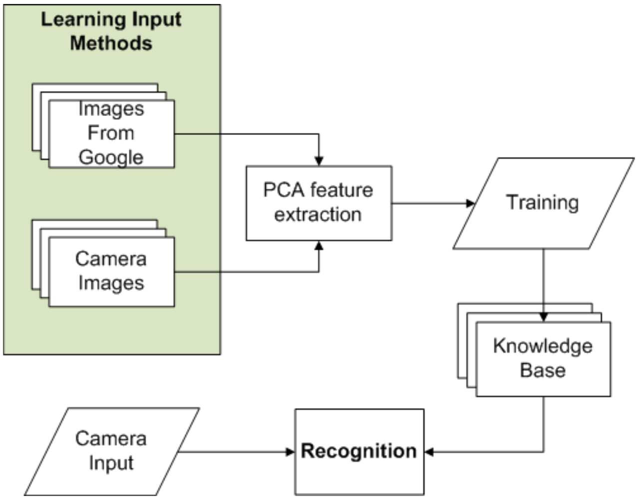

Fig. 2. The method we use to learn and recognize objects from web images. It is appreciated how training data-set is created and how one-class classification helps recognizing process.

3.1. Feature Extraction using PCA

For the image feature extraction Principal Component Analysis (PCA) is applied to the training images and feature vectors are obtained.

Let T be the training set containing the j images from one type of object, downloaded from Internet.

$$T = \{I_{T1}, I_{T2}, ..., I_{Tj}\}$$

The average image is calculated by:

$$I_{avg} = \frac{1}{j}\sum_{i=1}^{j} I_i \tag{1}$$

Now it is possible to apply PCA to the training images. PCA seeks the set of j orthonormal vectors u which best describes the distribution of the data. The vectors u_j and scalars λ_j are the eigenvectors and eigenvalues respectively, of the covariance matrix

$$\begin{aligned} C &= \tfrac{1}{j}\sum_{i=1}^{j} \Phi\Phi^T \\ &= AA^T \end{aligned} \tag{2}$$

Due to the image is a M by N matrix, the representing vector will be $M \times N$ length. We can keep the first $j-1$, eigenvalues and eigenvectors (from now on "eigenobjects") instead the $M \times N$ pixel positions in each image vector. This way, the number of variables is greatly reduced. For that, we are going to verify how useful each eigenvector is, so we calculate the percentage of accumulated information that the first $j-1$ eigenvectors provide:

$$E_{\%} = \frac{\sum_{i=1}^{j} \Phi_i}{\sum \Phi} \tag{3}$$

We can consider that $E_{\%} \geq 95\%$ is good enough to perform recognition and at this value, we let $j\prime = i$ and keep just the first $j\prime$ eigenobjects to represent images. Also, this way it is possible to keep some noise and irrelevant components out of the model. An example extracted from the our experiments is shown in figure 3.

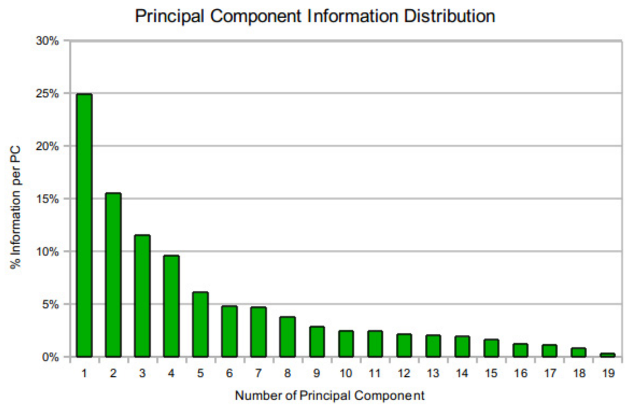

Fig. 3. Information percentage per Principal Component or Eigenvector from training for "mugs". This plot helps understand how the number of components can be reduced by the accumulated percentage.

Now, every image (I_{test}) to be classified is projected into the eigenspace by this operation:

$$\begin{gathered} \mathbf{u}_k^T = I_{test} - I_{avg} \\ k = 1, 2, \ldots j\prime \end{gathered} \tag{4}$$

where $\mathbf{u}_k^T$ represents the projected test vector to be compared against training set.

3.2. Recognition using k-NNDD

For classifying, first is necessary to segmentate objects in scene. In order to achieve this, the blob extraction functions provided by cvBlob [13, Carnero] were used. This way, it is possible to create a Region of Interest (ROI) and then extract it in order to project it over the eigenspace and classify it by k-NNDD.

When the test image is projected into the eigenspace, we apply the k-NNDD algorithm in order to decide if the image belongs to the objective class. To do so, we find the k nearest neighbors for the test image and apply this criterion:

$$f(z, NN_i(I_{test})) = \begin{cases} \text{Accept as target} & \frac{d(I_{test})}{d(NN_i(I_{test}))} \geq 1 \\ \text{Reject} & \frac{d(I_{test})}{d(NN_i(I_{test}))} < 1 \end{cases} \tag{5}$$

Where:

$$\frac{d(I_{test})}{d(NN_i(I_{test}))} = \frac{\sum_{i=1}^{k} \parallel NN_1(I_{test}) - NN_i(NN_1(I_{test})) \parallel}{\sum_{i=1}^{k} \parallel I_{test} - NN_i(I_{test}) \parallel} \tag{6}$$

4. Experiments and Results

This section presents the results of training the method on the robot. We used images from Internet belonging to mugs, bottles, books, phones and computer mice. We used training sets composed of 20 images of each object. Test sets where every 100 frames from the camera, in order to measure the method's performance.

As mentioned above, the scene on camera must be segmented. In order to do segmentation, we first convert the frame into grayscale, blur it to minimize the shadows and then use adaptive gaussian threshold to find edges. Once edges are detected, cvBlob library can find the ROI where objects are present. These ROI are extracted and passed to k-NNDD for recognition task. The segmentation process can be appreciated on figure 4.

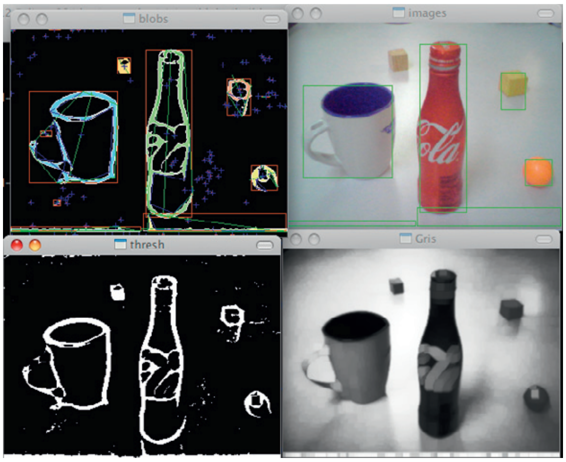

Fig. 4. Segmentation process. bottom-right: The grayscale-blurred frame. Bottom-left: Object edges. Top-left: The Blobs found. Top-right: ROI found in scene.

After found, ROI have to meet some aspect ratio restrictions. These are calculated from the training set and when a new ROI is found these restrictions are verified. Also, any ROI with *area* < 1500 are rejected, since they are too small to be taken into classification. Resolution used in experiments was 320×240 px. The output from the classifier for each of the considered objects is shown in figure 5.

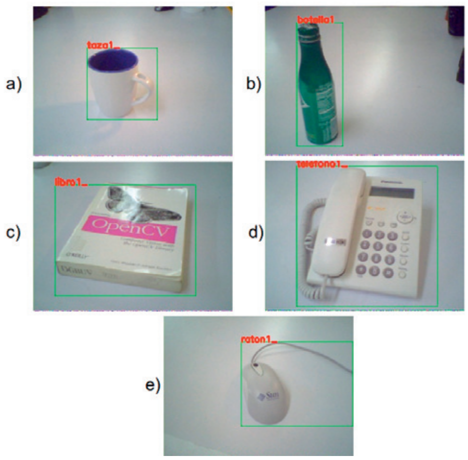

Fig. 5. The output from our method when training for: a) Mugs, b) Bottles, c) Books, d) Phones and, e) Computer Mouse.

In figure 6 it can be appreciated how our method only recognizes the objects it was trained for. In this example, the method was trained for mugs and only them are positive output. And it is noticeable that color is not a restriction, since this method is appearance-based.

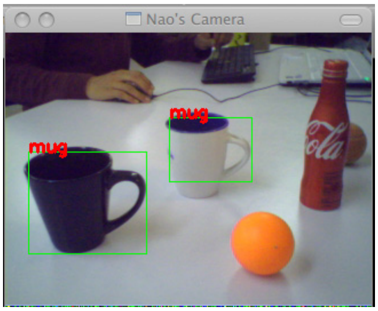

Fig. 6. The method trained for mugs, and with other objects in the scene. Only mugs are recognized.

The confusion matrix is presented in table 1. In the case of training for books, all telephones were misrecognized. This is because the shape of a book can easily been confused with the shape of the phone we use to test, as can be appreciated in figure 5. In case of training for telephones, there was not confusion, because phone components as buttons, wire, screen and others made a training set with more discriminating characteristics than books, giving a better performance.

Table 1. The confusion matrix from testing our method.

Test \ Train	**Mug**	**Bottle**	**Book**	**Phone**	**Mouse**
Mug	500	0	3	0	0
Bottle	0	500	0	0	0
Book	0	0	500	0	5
Phone	0	0	500	500	0
Mouse	0	0	0	0	493

Finally, the plot in figure 7 shows the average performance of the method we present. It is interesting to notice that a high performance is achieved even with a reduce number of images in the training sets.

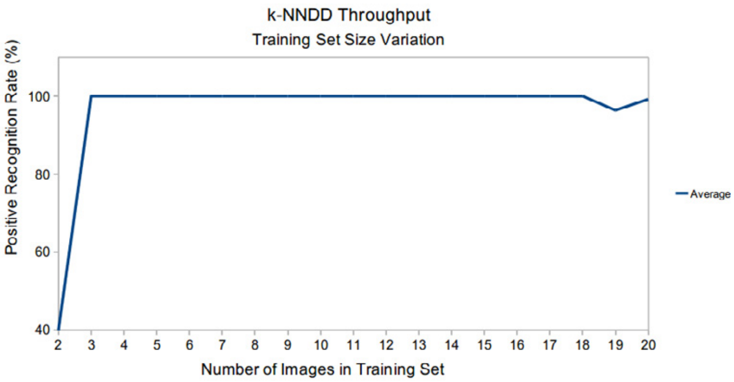

Fig. 7. Our method's average percentage performance for the whole set of objects tested

5. Conclusions and Future Work

We have presented in this paper, a method that provides to a service robot a flexible framework for learning and recognizing objects, thus without the needing of creating models for each one of them. As robot autonomy is an important subject in Human-Robot Interaction, using cloud resources such as Google's image searching engine and its huge storage capacity, becomes a viable option in order to save robot resources. The method also use PCA technique in order to reduce data dimensionality and then store the information needed to recognize the objects.

Also, we have presented the basis of a bigger architecture that presents different layers of data distribution. So the request for information is performed in four levels: First level data is locally retrieved from the robot itself, if robot does not have any information, the second level consists in collect information from other robots in the local network. If no partners available, in the third level data is acquired from the cloud, and finally from human by requesting the object to be shown.

The fact of finding the images we used for experiments over the Internet makes us conclude that it is possible to create the training set for many objects the robot needs to learn, but as we mentioned in Introduction section, ambiguity is a subject that has to be taken into account. In order to reduce it, we consider in future the integration of techniques like presented by Peñaloza *et. al.* [14] so in this way we can filter the downloaded images and keep only those which are more visually-related to the meaning of the keyword and those which show the real object.

References

[1] R. Arumugam, V. R. Enti, L. Bingbing, W. Xiaojun, K. Baskaran, F. F. Kong, A. S. Kumar, K. D. Meng, G. W. Kit, Davinci : A cloud computing framework for service robots, Architecture (2010) 3084–3089.
URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5509469`

[2] S. Kobayashi, Intelligent Humanoid Robot with Japanese Wikipedia Ontology and Robot Action Ontology, ACM, 2011, pp. 417–424.
URL `http://dl.acm.org/citation.cfm?id=1957656.1957811&coll=DL&dl=ACM&CFID=43206090&CFTOKEN=55604688`

[3] T. Cootes, C. Taylor, D. Cooper, J. Graham, et al., Active shape models—their training and application, Comput. Vis. Image Underst. 61 (1) (1995) 38–59. doi:10.1006/cviu.1995.1004.
URL `http://dx.doi.org/10.1006/cviu.1995.1004`

[4] H. Murase, S. Nayar, Visual Learning and Recognition of 3D Objects from Appearance, International Journal on Computer Vision 14 (1) (1995) 5–24.

[5] M. Turk, A. Pentland, M. I. of Technology. Media Laboratory. Vision, M. Group, Eigenfaces for Recognition, M.I.T. Media Lab Vision and Modeling Group technical report, Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.
URL `http://books.google.com.mx/books?id=frGOGwAACAAJ`

[6] M.A.Vicente, C. Fernández, R. Puerto, A. Gil, Extracción de características para reconocimiento visual, XXV Jornadas de Automática Ciudad Real, del 8 al 10 de septiembre de 2004.

[7] L. Malagón-Borja, O. Fuentes, Object detection using image reconstruction with pca, Image and Vision Computing 27 (1-2) (2009) 2–9.
URL `http://linkinghub.elsevier.com/retrieve/pii/S0262885607000820`

[8] K. Hempstalk, E. Frank, Discriminating against new classes: One-class versus multi-class classification, AI 2008 Advances in Artificial Intelligence 5360 (2008) 325–336.
URL `http://www.springerlink.com/index/82182j64595100u0.pdf`

[9] R. El-Yaniv, M. Nisenson, Optimal single-class classification strategies, Advances in Neural Information Processing Systems 19 41 (3-4) (2006) 377–384.
URL `http://eprints.pascal-network.org/archive/00002663/`

[10] V. D. Gesù, J. H. Friedman, G. L. Bosco, et al., Intruders pattern identification, 9th International Conference on Pattern Recognition.

[11] V. Laparra, J. Muñoz-Marí, G. Camps-Valls, J. Malo, Pca gaussianization for one-class remote sensing image classification, Image Processing Laboratory (IPL), Universitat de València.

[12] D. T. Munroe, M. G. Madden, Multi-class and single-class classification approaches to vehicle model recognition from images, Proceedings of IEEE AICS.

[13] C. C. Liñán, cvBlob, http://cvblob.googlecode.com.
URL `http://cvblob.googlecode.com`

[14] C. Peñaloza, Y. Mae, K. Ohara, T. Takubo, T. Arai, Web-based object category learning using human-robot interaction cues, Image Rochester NY (2011) 223–224.
URL `http://doi.acm.org/10.1145/1957656.1957742`